# *Chernoff Faces*

## Summary

**Chernoff Faces** provide a method for visualizing multivariate data by drawing cartoon faces in which various features are scaled according to the values of different quantitative variables. They were developed by Herman Chernoff and first described in the article titled "The Use of Faces to Represent Points in k-Dimensional Space Graphically", published in the Journal of the American Statistical Association, June 1973, Vol. 68, No. 342, pp. 361-368. While their effectiveness as a method for identifying groups of cases has been debated, they represent a novel alternative to more conventional multivariate visualization techniques.

## Sample StatFolio: *chernoff.sgp*

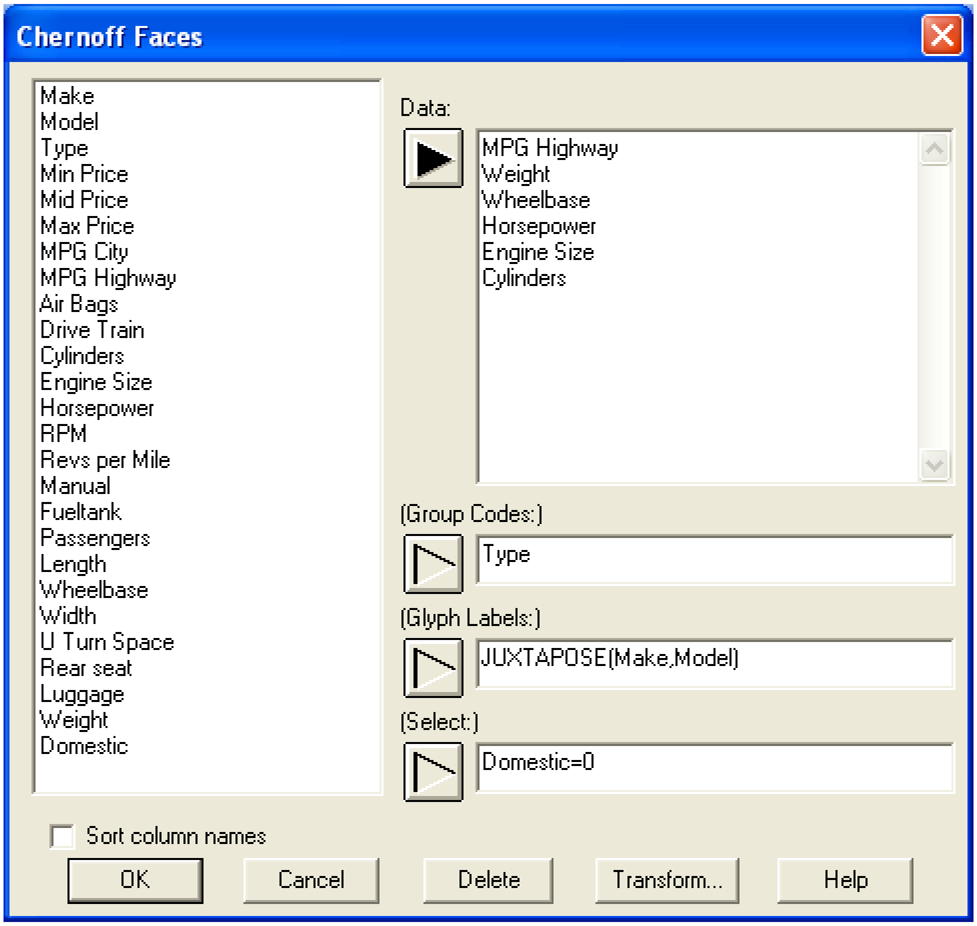
## Sample Data:

The file *93cars.sgd* contains information on 26 variables for *n* = 93 makes and models of automobiles, taken from Lock (1993). The table below shows a partial list of the data in that file:

| *Make* | *Model* | *MPG Highway* | *Weight* | *Wheelbase* | *Horsepower* | *Engine Size* | *Cylinders* |
|---|---|---|---|---|---|---|---|
| Acura | Integra | 31 | 2705 | 102 | 140 | 1.8 | 4 |
| Acura | Legend | 25 | 3560 | 115 | 200 | 3.2 | 6 |
| Audi | 90 | 26 | 3375 | 102 | 172 | 2.8 | 6 |
| Audi | 100 | 26 | 3405 | 106 | 172 | 2.8 | 6 |
| BMW | 535i | 30 | 3640 | 109 | 208 | 3.5 | 4 |
| Buick | Century | 31 | 2880 | 105 | 110 | 2.2 | 4 |
| Buick | LeSabre | 28 | 3470 | 111 | 170 | 3.8 | 6 |
| Buick | Roadmaster | 25 | 4105 | 116 | 180 | 5.7 | 6 |
| Buick | Riviera | 27 | 3495 | 108 | 170 | 3.8 | 6 |
| Cadillac | DeVille | 25 | 3620 | 114 | 200 | 4.9 | 8 |
| Cadillac | Seville | 25 | 3935 | 111 | 295 | 4.6 | 8 |
| Chevrolet | Cavalier | 36 | 2490 | 101 | 110 | 2.2 | 4 |
| Chevrolet | Corsica | 34 | 2785 | 103 | 110 | 2.2 | 4 |
| Chevrolet | Camaro | 28 | 3240 | 101 | 160 | 3.4 | 6 |
| Chevrolet | Lumina | 29 | 3195 | 108 | 110 | 2.2 | 4 |
| Chevrolet | Lumina_APV | 23 | 3715 | 110 | 170 | 3.8 | 6 |
| Chevrolet | Astro | 20 | 4025 | 111 | 165 | 4.3 | 6 |
| Chevrolet | Caprice | 26 | 3910 | 116 | 170 | 5.0 | 8 |
| Chevrolet | Corvette | 25 | 3380 | 96 | 300 | 5.7 | 8 |
| Chrylser | Concorde | 28 | 3515 | 113 | 153 | 3.3 | 6 |

## Data Input

The data to be analyzed consist of 2 or more numeric columns and an optional column with group identifiers:



Data:
MPG Highway
Weight
Wheelbase
Horsepower
Engine Size
Cylinders

(Group Codes:) Type

(Glyph Labels:) JUXTAPOSE(Make,Model)

(Select:) Domestic=0

- **Data:** 2 or more numeric columns containing the data to be plotted.

- **Group Codes:** an optional column with levels to be used to identify groups of cases.

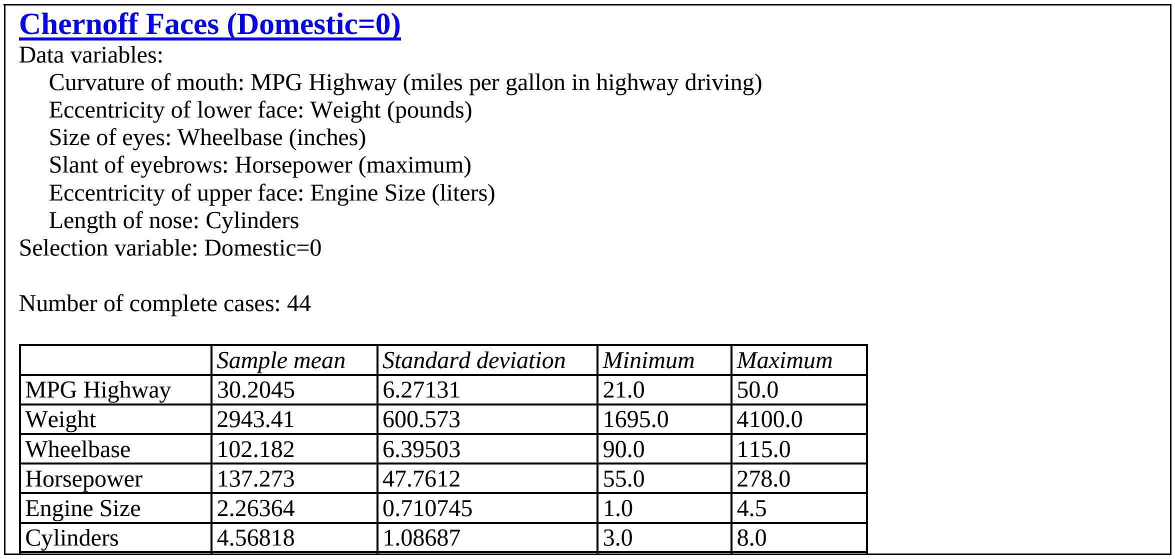
- **Glyph Labels:** an optional column with labels corresponding to each row. If not specified, row numbers will be used as labels.

- **Select:** subset selection.

As an example, 6 variables have been selected. The type of vehicle will be used to identify the cases. The JUXTAPOSE operator puts two columns side by side, so that each vehicle may be labeled with both its make and its model. The selection expression “Domestic = 0” specifies that only cars made outside the United States should be included.

## Analysis Summary

The *Analysis Summary* shows the number of rows with complete data and summary statistics for those rows:



**Chernoff Faces (Domestic=0)**

Data variables:
Curvature of mouth: MPG Highway (miles per gallon in highway driving)
Eccentricity of lower face: Weight (pounds)
Size of eyes: Wheelbase (inches)
Slant of eyebrows: Horsepower (maximum)
Eccentricity of upper face: Engine Size (liters)
Length of nose: Cylinders
Selection variable: Domestic=0

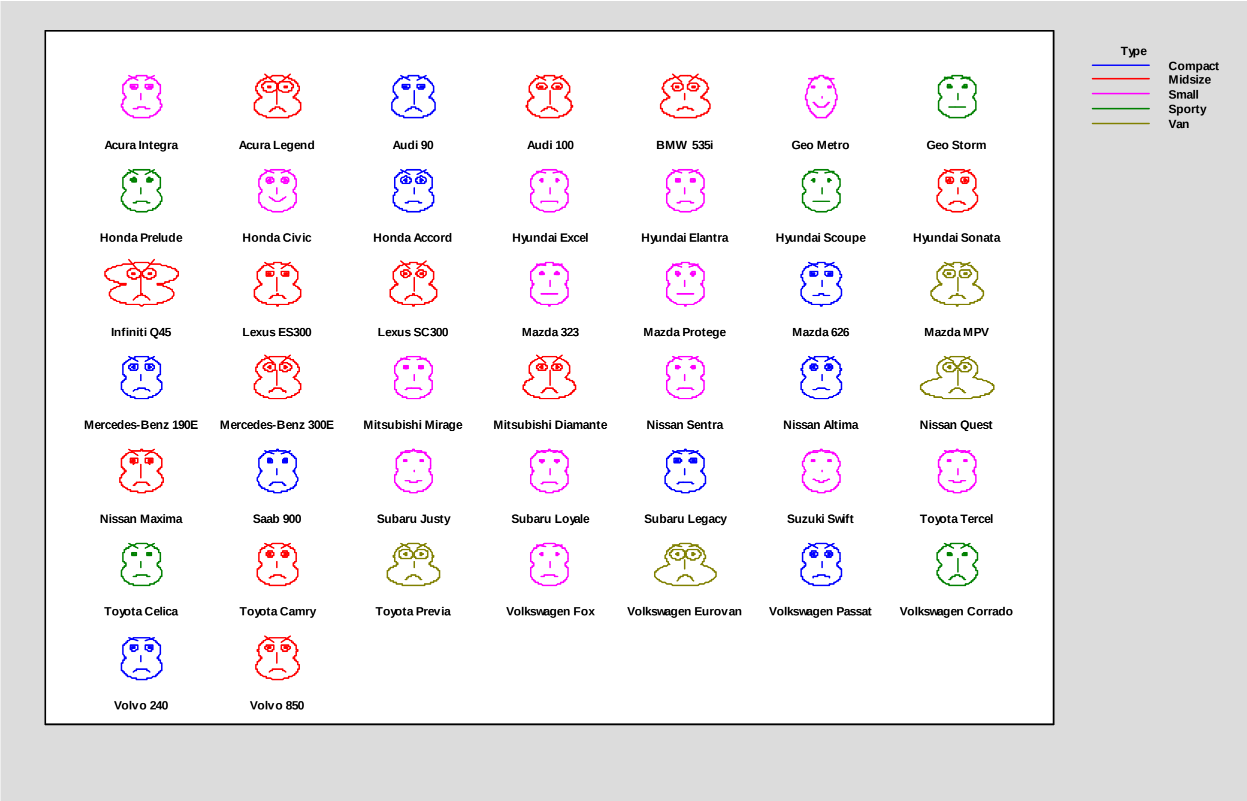
Number of complete cases: 44

| | *Sample mean* | *Standard deviation* | *Minimum* | *Maximum* |
|---|---|---|---|---|
| MPG Highway | 30.2045 | 6.27131 | 21.0 | 50.0 |
| Weight | 2943.41 | 600.573 | 1695.0 | 4100.0 |
| Wheelbase | 102.182 | 6.39503 | 90.0 | 115.0 |
| Horsepower | 137.273 | 47.7612 | 55.0 | 278.0 |
| Engine Size | 2.26364 | 0.710745 | 1.0 | 4.5 |
| Cylinders | 4.56818 | 1.08687 | 3.0 | 8.0 |

There are 44 rows which meet the selection criterion and have data for all of the variables. The output also shows which features of the face will be scaled according to each of the data variables.

## Chernoff Faces

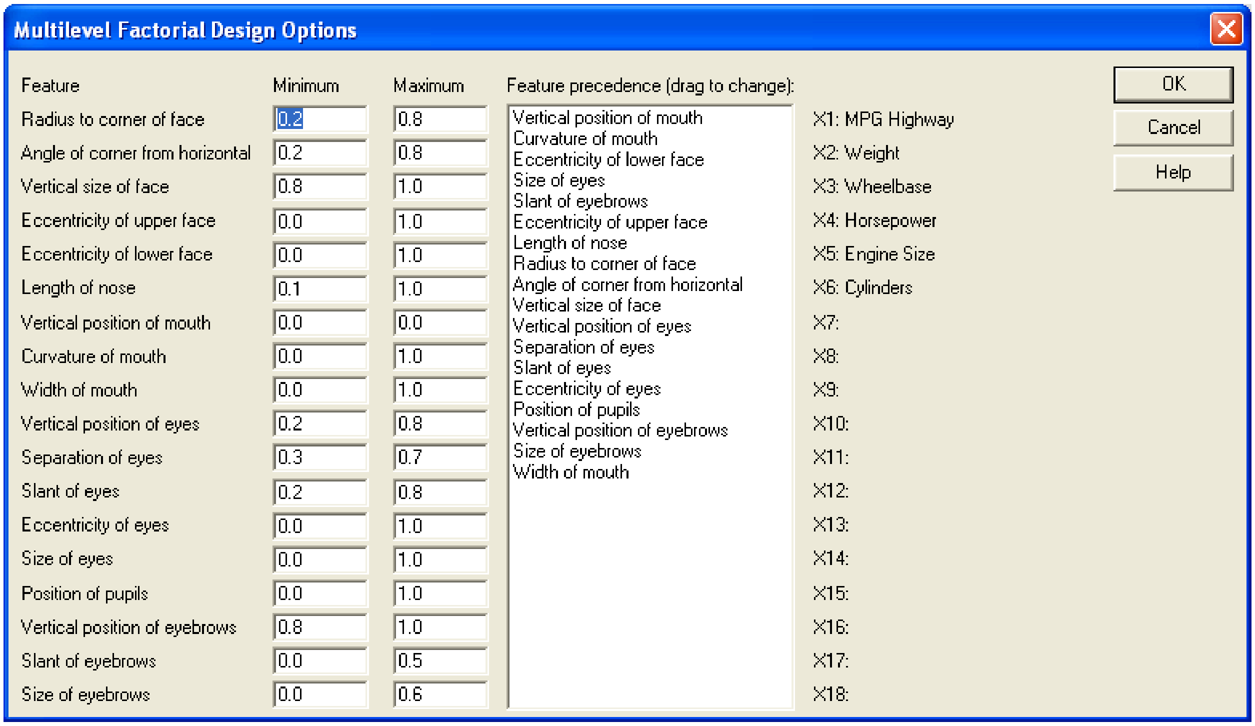
The display below shows a face for each observation in the selected data:



The color of the face indicates the type of vehicle. From this plot, you can identify vehicles with similar attributes (such as the Nissan Quest and the Volkswagen Eurovan) and also spot outliers (such as the Infiniti Q45).

## Analysis Options

The *Analysis Options* dialog box selects the features of the face that will be used to represent each variable:



| Feature | Minimum | Maximum |
|---|---|---|
| Radius to corner of face | 0.2 | 0.8 |
| Angle of corner from horizontal | 0.2 | 0.8 |
| Vertical size of face | 0.8 | 1.0 |
| Eccentricity of upper face | 0.0 | 1.0 |
| Eccentricity of lower face | 0.0 | 1.0 |
| Length of nose | 0.1 | 1.0 |
| Vertical position of mouth | 0.0 | 0.0 |
| Curvature of mouth | 0.0 | 1.0 |
| Width of mouth | 0.0 | 1.0 |
| Vertical position of eyes | 0.2 | 0.8 |
| Separation of eyes | 0.3 | 0.7 |
| Slant of eyes | 0.2 | 0.8 |
| Eccentricity of eyes | 0.0 | 1.0 |
| Size of eyes | 0.0 | 1.0 |
| Position of pupils | 0.0 | 1.0 |
| Vertical position of eyebrows | 0.8 | 1.0 |
| Slant of eyebrows | 0.0 | 0.5 |
| Size of eyebrows | 0.0 | 0.6 |

Feature precedence (drag to change):

| | |
|---|---|
| Vertical position of mouth | X1: MPG Highway |
| Curvature of mouth | X2: Weight |
| Eccentricity of lower face | X3: Wheelbase |
| Size of eyes | X4: Horsepower |
| Slant of eyebrows | X5: Engine Size |
| Eccentricity of upper face | X6: Cylinders |
| Length of nose | X7: |
| Radius to corner of face | X8: |
| Angle of corner from horizontal | X9: |
| Vertical size of face | X10: |
| Vertical position of eyes | X11: |
| Separation of eyes | X12: |
| Slant of eyes | X13: |
| Eccentricity of eyes | X14: |
| Position of pupils | X15: |
| Vertical position of eyebrows | X16: |
| Size of eyebrows | X17: |
| Width of mouth | X18: |

- **Minimum**: minimum scaling of each feature.

- **Maximum**: maximum scaling of each feature.

- **Feature precedence**: features assigned to each variable. The top feature in the list is assigned to X1, the second to X2, and so forth.

Each feature is scaled according to the value of a scaling parameter that ranges between 0 and 1. You may restrict the scaling to a narrower range if desired. The scaling of features not assigned to variables is set at a value halfway between the minimum and the maximum.

The faces are drawn by creating two overlapping ellipses, one higher than the other. The points at which the ellipses intersect are called the *corners* of the face. Each face is given a mouth, a nose, two eyes, and two eyebrows.

The scaling of each feature is defined by a value *h* that ranges between 0 and 1 according to

$$h = \frac{X - \min_x}{\max_x - \min_x} \qquad (1)$$

where $min_x$ and $max_x$ are the minimum and maximum observed values of the variable X. The details of each feature are described below:

1. *Radius to corner of face*: the distance from the center of the nose (point O) to the corners of the face (point P).

2. Angle of corner from horizontal: the angle from horizontal of a line drawn from O to P. The angle is defined such that the value 0.5 represents a situation where the corners of the faces are aligned with the center of the nose.

3. *Vertical size of face*: the vertical size of the face, where *h* represents half the vertical height.

4. *Eccentricity of upper face*: related to the ratio of the length of the major axis to that of the minor axis. *h* is scaled so that a value of 0.5 represents a circle. Values of *h* less than 0.5 correspond to ellipses in which the vertical axis is longer than the horizontal axis. Values of *h* greater than 0.5 correspond to ellipses in which the horizontal axis is longer than the vertical axis.

5. *Eccentricity of upper face*: the eccentricity of the lower face.

6. *Length of nose*: the length of the nose where *h* =1 equals half the vertical size of the face.

7. *Vertical position of mouth*: the position of the mouth. *h* =0 positions the mouth at a height equal to the center of the nose, while *h* = 1 positions the mouth at the bottom of the face.

8. *Curvature of mouth*: the amount of curvature in the mouth. If h = 0.5, the mouth has no curvature. For values of h < 0.5, the mouth forms a frown. For values of h > 0.5, the mouth forms a smile.

9. *Width of mouth*: the width of the mouth, where *h* = 1 corresponds to half the maximum width of the face.

10. *Vertical position of eyes*: the position of the eyes. *h* =0 positions the eyes at a height equal to the center of the nose, while *h* = 1 positions the eyes at the top of the face.

11. *Separation of eyes*: the distance between the center of the eyes, where *h* = 1 corresponds to half the maximum width of the face.

12. *Slant of eyes*: the amount by which the eyes slant. If h = 0.5, the eyes have no slant. For values of h < 0.5, the eyes slant one way. For values of h > 0.5, the eyes slant the other way.

13. *Eccentricity of eyes*: related to the ratio of the length of the major axis of the ellipse that forms each eye to that of the minor axis. *h* is scaled so that a value of 0.5 represents a circle. Values of *h* less than 0.5 correspond to ellipses in which the vertical axis is longer than the horizontal axis. Values of *h* greater than 0.5 correspond to ellipses in which the horizontal axis is longer than the vertical axis.

14. *Size of eyes*: the width of the eyes.

15. *Position of pupils*: the position of the pupils within the eyes, where *h* = 0.5 puts the pupils in the middle of the eyes.

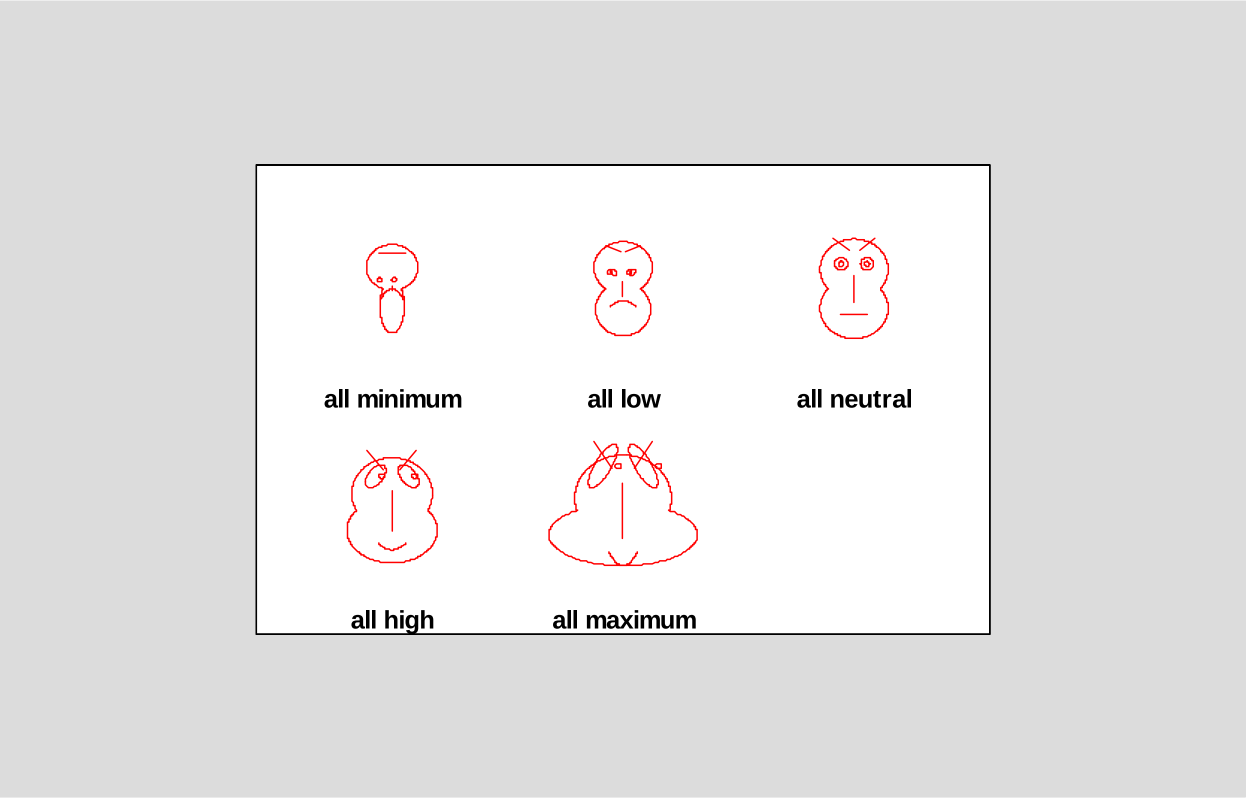
16. *Vertical position of eyebrows*: the position of the eyebrows. *h* =0 positions the eyebrows at a height equal to the center of the nose, while *h* = 1 positions the eyebrows at the top of the face.

17. *Slant of eyebrows*: the amount by which the eyebrows slant. If h = 0.5, the eyebrows have no slant. For values of h < 0.5, the eyebrows slant one way. For values of h > 0.5, the eyebrows slant the other way.

18. *Size of eyebrows*: the width of the eyebrows.

## *Key Glyph*

The *Key Glyph* dialog box shows you how the faces will look at selected combinations of the variables:



The combinations are:

1. *All minimum* – all variables are set at the minimum values observed in the data.
2. *All low* - all variables are set halfway between the minimum values and the midrange.
3. *All neutral* - all variables are set at the midrange.
4. *All high* - all variables are set halfway between the midrange and the maximum value.
5. *All maximum* – all variables are set at the maximum values observed in the data.